Learning From Time

Daniela Witten
Departments of Statistics & Biostatistics
University of Washington

June 9, 2016
Collaborators

Ali Shojaei

Shizhe Chen
Graphical Model
Graphical Model
Graphical Model
Goal: Learn the Structure of the Graph
Goal: Learn the Structure of the Graph
Time is Important

mesodermal progenitor → myoblasts → multinucleate myotube → muscle fiber

http://www.mun.ca/biology/desmid/brian/BIOL3530/DEV010/devo10.html
Time is Important
Time is Important

mesodermal progenitor → myoblasts → multinucleate myotube → muscle fiber

4 beakers with green liquid
Time is Important

- mesodermal progenitor → myoblasts → multinucleate myotube → muscle fiber
Time is Important

mesodermal progenitor → myoblasts → multinucleate myotube → muscle fiber

Tubes with green liquid indicate the progression of time.
Time is Important
Time is Important
Part I: Learning Gene Regulatory Relationships
Gene Expression Data
Gene Expression Data
Gene Expression Data
Multivariate Time-Course Data
Multivariate Time-Course Data
Multivariate Time-Course Data
Multivariate Time-Course Data
Multivariate Time-Course Data
Noiseless Trajectories

![Graph showing noiseless trajectories with time on the x-axis and value on the y-axis. The trajectories are color-coded and show different behaviors over time.](image)
Noiseless Trajectories
Noiseless Trajectories
Noiseless Trajectories
Noiseless Trajectories

\[ Y_j(t_i) = X_j(t_i) + \epsilon_j(t_i) \]
Noiseless Trajectories

\[ Y_j(t_i) = X_j(t_i) + \epsilon_j(t_i) \]
Noiseless Trajectories

$$Y_j(t_i) = X_j(t_i) + \epsilon_j(t_i)$$
A Model for the Noiseless Trajectories

For $j = 1, \ldots, p$,

$$\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),$$

where $f_{jk}$ is unknown.
A Model for the Noiseless Trajectories

For \( j = 1, \ldots, p \),

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),
\]

where \( f_{jk} \) is unknown.

\[
\begin{align*}
\frac{d}{dt} X_1(t) &= X_2^2(t) + \exp(X_2(t)) \\
\frac{d}{dt} X_2(t) &= 1 + \log(X_3(t)) \\
\frac{d}{dt} X_3(t) &= 2
\end{align*}
\]
A Model for the Noiseless Trajectories

For \( j = 1, \ldots, p \),

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)),
\]

where \( f_{jk} \) is unknown.

\[
\begin{align*}
\frac{d}{dt} X_1(t) &= X_2^2(t) + \exp(X_2(t)) \\
\frac{d}{dt} X_2(t) &= 1 + \log(X_3(t)) \\
\frac{d}{dt} X_3(t) &= 2
\end{align*}
\]
Challenges in Fitting the Model, Part I

\[ \frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{P} f_{jk}(X_k(t)) \]

Ravikumar et al. (2009)
Challenges in Fitting the Model, Part I

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{P} f_{jk}(X_k(t))
\]

Challenge: \( f_{jk}(\cdot) \) is unknown.

Ravikumar et al. (2009)
Challenges in Fitting the Model, Part I

The differential equation is given by:

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t))
\]

**Challenge:** \( f_{jk}(\cdot) \) is unknown.

**Solution:** Approximate with basis functions, \( \psi_1(\cdot), \ldots, \psi_M(\cdot) \):

\[
\frac{d}{dt} X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^T \theta_{jk}
\]

Ravikumar et al. (2009)
Challenges in Fitting the Model, Part II

\[ \frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^T \theta_{jk} \]

Solution: Group lasso approach to induce sparsity.

Yuan and Lin (2006); Simon and Tibshirani (2012)
Challenges in Fitting the Model, Part II

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk} (X_k(t)) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^T \theta_{jk}
\]

Challenge: \( O(Mp^2) \) unknown parameters and \( N \) timepoints.

Yuan and Lin (2006); Simon and Tibshirani (2012)
Challenges in Fitting the Model, Part II

\[
\frac{d}{dt} X_j(t) = C_j + \sum_{k=1}^{p} f_{jk}(X_k(t)) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t))^T \theta_{jk}
\]

Challenge: \(O(Mp^2)\) unknown parameters and \(N\) timepoints.

Solution: Group lasso approach to induce sparsity.

Yuan and Lin (2006); Simon and Tibshirani (2012)
Challenges in Fitting the Model, Part III

\[
\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{P} f_{jk}(X_k(t)) \approx C_j + \sum_{k=1}^{P} \psi(X_k(t))^T \theta_{jk}
\]
Challenges in Fitting the Model, Part III

\[ \frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{P} f_{jk}(X_k(t)) \approx C_j + \sum_{k=1}^{P} \psi(X_k(t))^T \theta_{jk} \]

Challenge: \( X_k(t) \) is unobserved.
Challenges in Fitting the Model, Part III

\[
\frac{d}{dt}X_j(t) = C_j + \sum_{k=1}^{P} f_{jk}(X_k(t)) \approx C_j + \sum_{k=1}^{P} \psi(X_k(t))^T \theta_{jk}
\]

Challenge: \(X_k(t)\) is unobserved.

Solution: Estimate \(X_k(t)\) using \(Y_k(t_1), \ldots, Y_k(t_N)\).
Existing Methods Estimate the Derivative

\[
\frac{d}{dt} X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}
\]

---

Wu et al. (2014) and Henderson and Michailidis (2014)
Existing Methods Estimate the Derivative

\[
\frac{d}{dt} X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}
\]
Existing Methods Estimate the Derivative

\[
\frac{d}{dt} X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(\hat{X}_k(t)) \cdot \theta_{jk}
\]
Existing Methods Estimate the Derivative

\[ \frac{d}{dt} \hat{X}_j(t) \approx C_j + \sum_{k=1}^{p} \psi(\hat{X}_k(t)) \cdot \theta_{jk} \]
Estimating the Derivative is Hard

\[
\frac{d}{dt} \hat{X}_j(t) \text{ and } \frac{d}{dt} X_j(t)
\]
Instead, We Can Integrate

\[
\frac{d}{dt} X_j(t) \approx C_j + \sum_{k=1}^{p} \psi(X_k(t)) \cdot \theta_{jk}
\]

The idea of integrating is due to Dattner and Klaassen (2013)
Instead, We Can Integrate

\[
\int_0^{t_i} \frac{d}{dt} X_j(s) ds \approx \int_0^{t_i} C_j \, ds + \int_0^{t_i} \sum_{k=1}^p \psi(X_k(s)) \cdot \theta_{jk} \, ds
\]

The idea of integrating is due to Dattner and Klaassen (2013)
Instead, We Can Integrate

The idea of integrating is due to Dattner and Klaassen (2013)

\[ X_j(t_i) - X_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[ \int_{0}^{t_i} \psi(X_k(s))ds \right] \cdot \theta_{jk} \]
Instead, We Can Integrate

\[ Y_j(t_i) - X_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[ \int_{0}^{t_i} \psi(X_k(s)) ds \right] \cdot \theta_{jk} \]

The idea of integrating is due to Dattner and Klaassen (2013)
Instead, We Can Integrate

\[ Y_j(t_i) - \hat{X}_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[ \int_{0}^{t_i} \psi(X_k(s)) ds \right] \cdot \theta_{jk} \]

The idea of integrating is due to Dattner and Klaassen (2013)
Instead, We Can Integrate

\[ Y_j(t_i) - \hat{X}_j(0) \approx t_i C_j + \sum_{k=1}^{p} \left[ \int_{0}^{t_i} \psi(\hat{X}_k(s)) \, ds \right] \cdot \theta_{jk} \]
Estimating the Integral is Easy

\[
\int_0^{t_i} \psi(\hat{X}_k(s)) \, dt \quad \text{and} \quad \int_0^{t_i} \psi(X_k(s)) \, ds
\]
Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve
$$\minimize_{Z(\cdot) \in \chi(h)} \left\{ \sum_{i=1}^{n} \|Y_j(t_i) - Z(t_i)\|_2 \right\}.$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_j^1, \ldots, \hat{\theta}_j^p \in \mathbb{R}^M$ that minimize
$$\int \|d\frac{dt}{dt} \hat{X}_j(t) - C_j - p \sum_{k=1}^{\psi} \psi(\hat{X}_k(t))^T \theta_{jk} \|_2^2 dt + \lambda p \sum_{k=1}^{\psi} \int \psi(\hat{X}_k(t))^T \theta_{jk} \|_2^2 dt.$$

Step 3: The graph estimate is $\hat{E} = \{ (j, k) : \hat{\theta}_{jk} \neq 0 \}$.

Wu et al. (2014) and Henderson and Michailidis (2014)
Existing Methods Estimate the Derivative

Step 1: For \( j = 1, \ldots, p \), let \( \hat{X}_j(\cdot) \) solve

\[
\minimize_{Z(\cdot) \in \mathcal{X}(h)} \left\{ \sum_{i=1}^{n} \left\| Y_j(t_i) - Z(t_i) \right\|^2 \right\}.
\]

Step 2: For \( j = 1, \ldots, p \), find \( \hat{\theta}_j^1, \ldots, \hat{\theta}_j^p \in \mathbb{R}^M \) that minimize

\[
\begin{aligned}
&\int dt \left\| \frac{d}{dt} \hat{X}_j(t) - C_j - p \sum_{k=1}^{p} \psi(\hat{X}_k(t))^{T} \theta_{jk} \right\|^2 \\
&+ \lambda p \sum_{k=1}^{p} \sqrt{\int \left( \psi(\hat{X}_k(t))^{T} \theta_{jk} \right)^2 dt}
\end{aligned}
\]

standardized group lasso.

Step 3: The graph estimate is \( \hat{E} = \{ (j, k) : \hat{\theta}_{jk} \neq 0 \} \).

---

Wu et al. (2014) and Henderson and Michailidis (2014)
Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$
\minimize_{Z(\cdot) \in \chi(h)} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.
$$

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$
\int \left\| \frac{d}{dt} \hat{X}_j(t) - C_j - \sum_{k=1}^{p} \psi(\hat{X}_k(t))^T \theta_{jk} \right\|^2 dt
+ \lambda \sum_{k=1}^{p} \sqrt{\int \left( \psi(\hat{X}_k(t))^T \theta_{jk} \right)^2 dt}.
$$

standardized group lasso

Wu et al. (2014) and Henderson and Michailidis (2014)
Existing Methods Estimate the Derivative

Step 1: For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\minimize_{Z(\cdot) \in \mathcal{X}(h)} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$ 

Step 2: For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\int \left\| \frac{d}{dt} \hat{X}_j(t) - C_j - \sum_{k=1}^{p} \psi(\hat{X}_k(t))^T \theta_{jk} \right\|^2 dt +$$

$$+ \lambda \sum_{k=1}^{p} \sqrt{\int \left( \psi(\hat{X}_k(t))^T \theta_{jk} \right)^2} dt.$$ 

standardized group lasso

Step 3: The graph estimate is $\hat{E} = \{(j, k) : \hat{\theta}_{jk} \neq 0\}$.

Wu et al. (2014) and Henderson and Michailidis (2014)
Our Proposal:

**Graph Reconstruction w/ Additive Differential Equations**

Step 1: For \( j = 1, \ldots, p \), let \( \hat{X}_j(\cdot) \) solve

\[
\text{minimize } \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|_2^n
\]

Step 2: For \( j = 1, \ldots, p \), find \( \hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M \) that minimize

\[
\sum_{i=1}^{n} \left[ Y_j(t_i) - \hat{X}_j(0) - t_i C_j - \sum_{k=1}^{p} \hat{\Psi}_T ik \theta_{jk} \right]^2 + \lambda \sum_{k=1}^{p} \left( \sum_{i=1}^{n} (\hat{\Psi}_T ik \theta_{jk})^2 \right)^{1/2}
\]

where \( \hat{\Psi}_{ik} = \int_{t_i}^0 \psi(\hat{X}_k(s)) ds \), \( i = 1, \ldots, n \).

Step 3: The graph estimate is \( \hat{E} = \{ (j, k) : \hat{\theta}_{jk} \neq 0 \} \).
Our Proposal:

**Graph Reconstruction w/ Additive Differential Equations**

**Step 1:** For \( j = 1, \ldots, p \), let \( \hat{X}_j(\cdot) \) solve

\[
\minimize_{Z(\cdot) \in \chi(h)} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.
\]
Our Proposal: 

**Graph Reconstruction w/ Additive Differential Equations**

**Step 1:** For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$\minimize_{Z(\cdot) \in \chi(h)} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.$$

**Step 2:** For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$\sum_{i=1}^{n} \left[ Y_j(t_i) - \hat{X}_j(0) - t_i C_j - \sum_{k=1}^{p} \hat{\Psi}_{ik}^T \hat{\theta}_{jk} \right]^2 + \lambda \sqrt{\sum_{i=1}^{n} \left( \hat{\Psi}_{ik}^T \hat{\theta}_{jk} \right)^2},$$

where $\hat{\Psi}_{ik} = \int_{0}^{t_i} \psi(\hat{X}_k(s)) \, ds$, $i = 1, \ldots, n$. 

standardized group lasso
Our Proposal:

**Graph Reconstruction w/ Additive Differential Equations**

**Step 1:** For $j = 1, \ldots, p$, let $\hat{X}_j(\cdot)$ solve

$$
\minimize_{Z(\cdot) \in \chi(h)} \left\{ \sum_{i=1}^{n} \| Y_j(t_i) - Z(t_i) \|^2 \right\}.
$$

**Step 2:** For $j = 1, \ldots, p$, find $\hat{\theta}_{j1}, \ldots, \hat{\theta}_{jp} \in \mathbb{R}^M$ that minimize

$$
\sum_{i=1}^{n} \left[ Y_j(t_i) - \hat{X}_j(0) - t_i C_j - \sum_{k=1}^{p} \hat{\Psi}_{ik}^T \hat{\theta}_{jk} \right]^2 + \lambda \sum_{k=1}^{p} \sqrt{\sum_{i=1}^{n} \left( \hat{\Psi}_{ik}^T \hat{\theta}_{jk} \right)^2},
$$

where $\hat{\Psi}_{ik} = \int_{0}^{t_i} \psi(\hat{X}_k(s)) \, ds$, $i = 1, \ldots, n$.

**Step 3:** The graph estimate is $\hat{E} = \left\{ (j, k) : \hat{\theta}_{jk} \neq 0 \right\}$. 
Theory – Overview Of Our Results

- We bound
  \[ \int_t \left\{ \hat{X}_j(t) - X_j(t) \right\}^2 dt, \]
  which allows us to bound \( \|\hat{\Psi} - \Psi\| \) in high dimensions.

---

Extending Loh and Wainwright (2012)
We bound

\[ \int_t \left\{ \hat{X}_j(t) - X_j(t) \right\}^2 dt, \]

which allows us to bound \( \|\hat{\Psi} - \Psi\| \) in high dimensions.

We establish variable selection consistency of (standardized) group lasso regression with errors-in-variables.
Theory – Overview Of Our Results

- We bound

\[ \int_t \left\{ \hat{X}_j(t) - X_j(t) \right\}^2 dt, \]

which allows us to bound \( \| \hat{\Psi} - \Psi \| \) in high dimensions.
- We establish variable selection consistency of (standardized) group lasso regression with errors-in-variables.
- We show that with high probability, GRADE correctly identifies the parents of each node.

Extending Loh and Wainwright (2012)
Simulation Results

- **NeRDS**: Network Reconstruction via Dynamic Systems
- **GRADE**

NeRDS is the proposal of Henderson and Michailidis (2014)
The End Result
The End Result

![Graph showing Value over Time with a network diagram on the right]
Part II: Learning Functional Connectivity Among Neurons
Neuronal Spike Train Data

See e.g. Pillow et al. (2008)
Neuronal Spike Train Data
Neuronal Spike Train Data
Neuronal Spike Train Data
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
The Hawkes Process

Hawkes (1971)
Goal
The Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i}) \]

- \( \lambda_j(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R} \): intensity function
- \( \mu_j \in \mathbb{R} \): background intensity
- \( \omega_{j,k}(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R} \): transfer function
- \( t_{k,i} \in \mathbb{R}^+ \): time at which the \( k \)th neuron has its \( i \)th spike

Hawkes (1971)
Graph Corresponding to the Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i}) \]
Graph Corresponding to the Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_k, i \leq t} \omega_{j,k}(t - t_{k,i}) \]

\[ \omega_{1,2}(t) \neq 0 \]
\[ \omega_{2,3}(t) \neq 0 \]
Graph Corresponding to the Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i}) \]

\[ \omega_{1,2}(t) \neq 0 \]

\[ \omega_{2,3}(t) \neq 0 \]
Challenges in Fitting the Model, Part I

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i:t_k,i \leq t} \omega_{j,k}(t - t_{k,i}) \]
Challenges in Fitting the Model, Part I

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: \tau_k,i \leq t} \omega_{j,k}(t - \tau_{k,i}) \]

Challenge: The transfer function \( \omega_{j,k}(\cdot) \) is unknown.
\[
\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_k,i \leq t} \omega_{j,k}(t - t_k,i)
\]

**Challenge:** The transfer function \( \omega_{j,k}(\cdot) \) is unknown.

**Solution:** Approximate with basis functions, \( \psi_1(\cdot), \ldots, \psi_M(\cdot) \):

\[
\lambda_j(t) \approx \mu_j + \sum_{k=1}^{p} \sum_{i: t_k,i \leq t} [\psi(t - t_k,i)]^T \beta_{jk}
\]
Challenges in Fitting the Model, Part II

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_k,i \leq t} \omega_{j,k}(t - t_{k,i}) \]
Challenges in Fitting the Model, Part II

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_k, i \leq t} \omega_{j,k}(t - t_{k,i}) \]

Challenge: Need to estimate \( p^2 \) transfer functions, where \( p \) is large.
\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_k, i \leq t} \omega_{j,k}(t - t_{k,i}) \]

**Challenge:** Need to estimate \( p^2 \) transfer functions, where \( p \) is large.

**Solution:** Group lasso to induce sparsity in transfer functions.
Our Proposal: Neighborhood Selection Approach

Step 1: For $j = 1,...,p$, find $\hat{\beta}_j^1,...,\hat{\beta}_j^p \in \mathbb{R}^M$ that minimize $L_j(\beta_j^1,...,\beta_j^p) + \lambda \sum_{k=1}^p \|\psi^T \beta_j^k, k\|_2$.

Step 2: The graph estimate is $\hat{E} = \{(j, k): \hat{\beta}_{jk} \neq 0\}$.

Related Work: Meinshausen and Bühlmann (2006); Zhou et al. (2013a,b); Bacry et al. (2015); Hansen et al. (2015)
Our Proposal: Neighborhood Selection Approach

Step 1: For $j = 1, \ldots, p$, find $\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jp} \in \mathbb{R}^M$ that minimize

$$L_j(\beta_{j1}, \ldots, \beta_{jp}) + \lambda \sum_{k=1}^{p} \|\psi^T \beta_{j,k}\|_2.$$
Our Proposal: Neighborhood Selection Approach

**Step 1:** For $j = 1, \ldots, p$, find $\hat{\beta}_{j1}, \ldots, \hat{\beta}_{jp} \in \mathbb{R}^M$ that minimize

$$L_j(\beta_{j1}, \ldots, \beta_{jp}) + \lambda \sum_{k=1}^{p} \|\psi^T \beta_{j,k}\|_2.$$ 

**Step 2:** The graph estimate is $\hat{\mathcal{E}} = \{(j, k) : \hat{\beta}_{jk} \neq 0\}$.

---

Related Work: Meinshausen and Bühlmann (2006); Zhou et al. (2013a,b); Bacry et al. (2015); Hansen et al. (2015)
Theoretical Results

We establish model selection consistency in high dimensions; i.e. the parent set of each neuron is correctly estimated.
The End Result
The End Result
The End Result
Summary of Pipeline

Data
Summary of Pipeline

Data → Model
Summary of Pipeline
Summary of Pipeline

Data → Model → Regularize → Model Selection Consistency
Summary of Pipeline
Model Selection Consistency
Model Selection Consistency
Model Selection Consistency
Model Selection Consistency

As the number of timepoints grows, this is unlikely to happen.
Model Selection Consistency

“Zeroth-Order Inference”
What Does First-Order Inference Look Like?
What Does First-Order Inference Look Like?
What Does First-Order Inference Look Like?
What Does First-Order Inference Look Like?

What Does First-Order Inference Look Like?

- **P-value** associated with each edge?

---

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?
- Posterior distribution?

What Does First-Order Inference Look Like?

- P-value associated with each edge?
- False discovery rate associated with the estimated edge set?
- Posterior distribution?

Model certainly does not hold!

Summary

▶ Learn graph structure from temporal data.
▶ Different data, different models.
▶ Common themes:
  ▶ Do not assume functional form: use basis expansions.
  ▶ Estimate a sparse graph using group lasso penalties.
  ▶ Establish that the estimated graph is correct w.h.p.
Learn graph structure from temporal data.
Summary

- Learn graph structure from temporal data.
- Different data, different models.
Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
  - Estimate a sparse graph using group lasso penalties.
  - Establish that the estimated graph is correct w.h.p.
Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
  - Estimate a sparse graph using group lasso penalties.
Summary

- Learn graph structure from temporal data.
- Different data, different models.
- Common themes:
  - Do not assume functional form: use basis expansions.
  - Estimate a sparse graph using group lasso penalties.
  - Establish that the estimated graph is correct w.h.p.
What is Missing?

Model and assumptions certainly do not hold. Now what?

W.h.p. the estimated graph is 100% correct — but if not, all bets are off.

Can I get a p-value for each edge, or a false discovery rate?

Do I really believe the estimated graph?

Next steps for a biological collaborator?

No gold standard.
What is Missing?

- Model and assumptions certainly do not hold. Now what?
What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?
What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?
- Do I really believe the estimated graph?
What is Missing?

- Model and assumptions certainly do not hold. Now what?
- W.h.p. the estimated graph is 100% correct — but if not, all bets are off.
- Can I get a p-value for each edge, or a false discovery rate?
- Do I really believe the estimated graph?
  - Next steps for a biological collaborator?
  - No gold standard.


